Research & Methods
ISSN 1234-9224 Vol. 22 (1, 2013): 55–77
Institute of Philosophy and Sociology
Polish Academy of Sciences, Warsaw
www.ifispan.waw.pl
e-mail: publish@ifispan.waw.pl

How (Not) to Estimate the Design Effect of a Complex Sampling Scheme: A Case Study of the Polish Section of the European Social Survey, Round 5

Piotr Jabkowski
Institute of Sociology, Adam Mickiewicz University in Poznan, Poland

Design effect (DEFF) is a measure used to assess the effectiveness of a particular sampling scheme. Even though its definition is remarkably simple (cf. Kish 1965: 258), its practical implementation turns out to be problematic. Researchers therefore usually simplify the estimation of DEFF by independently determining the values of three components, namely, the clustering effect (DEFF*c*), the stratification effect (DEFF*s*) and the effect of unequal sampling probabilities (DEFF*p*) and by multiplying these partial measures to obtain a measure of overall effect. However, the validity of such a simplified version depends on strict formal requirements which are met only in a few sampling schemes. The subject of the analysis presented here is the sampling scheme in the Polish section of round 5of the European Social Survey (ESS). It will be shown that the method of DEFF estimation applied by the Polish coordinators of the project, which is compatible with the methodological recommendations of ESS (cf. Lynn et al. 2007: 114),does not satisfy the formal criteria that would validate its use. The author proposes two other ways of estimating the size of DEFF (cf. Gabler et al. 2006: 116-117) appropriate for the sampling scheme in ESS5-PL. Empirical analyses indicate that the use of the simplified procedure of DEFF prediction leads to significant underestimation of variance inflation in the sample design of ESS5-PL and, in turn, to overestimation of effective sample size.

Key words: European Social Survey; design effect; complex sampling; clustering effect; stratification effect; effect of unequal sampling probabilities

DESIGN EFFECT AND ITS COMPLICATIONS

It is a well-known fact that the variance of estimators is conditioned not only by the size of a sample or the level of diversity of individuals in a population, but also, or mostly, by all the activities undertaken in the course of selecting samples in survey research. In the monograph *Survey Errors and Survey Costs*, Groves (1989: 252) specifies three main types of sampling schemes which affect the precision of measurement:

> "The sample design features which are most important in this regard are (1) stratification, the sorting of the population into separate groups prior to selection, (2) assignment of probabilities of selection to different kinds of elements in the population, and (3) clustering, the selection of groups of elements together instead of independent selection of separate elements" (Groves 1989: 252).

In other words, if a researcher is not able to select a sample according to the rule of simple random sampling, then each applied sample design (of the same size as a simple random sample) results in a loss (or, occasionally, an increase) in the precision of estimation (Dorofeev et al. 2006: 4). The measure of design effect(DEFF[1]) is defined as the "ratio of the actual variance of a sample to the variance of a simple random sample of the same number of elements" (Kish 1965: 258).

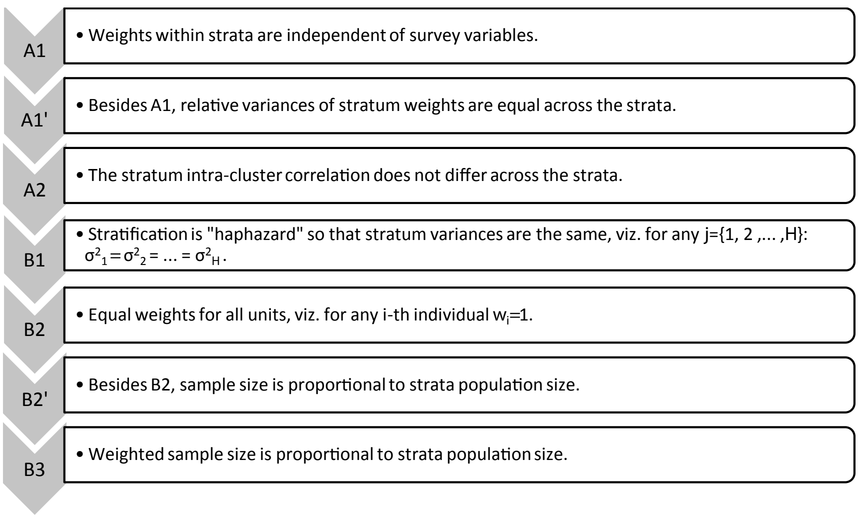
The definition of DEFF presented here is very simple, yet its practical implementation is a difficult task. In multi-stage and complex sampling, estimators of population parameters take the form of quite complicated mathematical formulae. Of course, the major complication does not arise from the complexity of the calculation, but from the need to devise methods of variance estimation for these estimators which would be appropriate for sample design. There is no ideal solution to this problem, but many researchers today simplify DEFF estimation by independently estimating the effects of stratified sampling ($DEFF_s$), cluster sampling ($DEFF_c$) and/or selection with unequal sampling probabilities ($DEFF_p$), and then multiplying these single indicators to obtain a measure of overall effect. Kish (1987: 202) and others[2] define DEFF as:

(1) $$DEFF_{TOTAL} = DEFF_p \times DEFF_s \times DEFF_c \text{, where:}$$

$DEFF_p$ is a measure of variance inflation that is obtained from the unequal probabilities of selection of individuals in a sample (Lynn et al. 2007: 112-113, Gabler et al. 1999: 105), $DEFF_s$ is a measure of the effect of selecting a sample from stratified populations (Dorofeev et al. 2006: 94), and $DEFF_c$ is a measure of the effect of cluster sampling, which according to Kish (1965: 162) is determined by intra-cluster correlation coefficient[3] (known as ρ). Then $DEFF_c = 1 + (b' - 1)\rho$, where b' is a parameter representing the size of the clusters.

Although this estimation of $DEFF_{TOTAL}$ is widely used today, even in the research on ESS (Lynn et al. 2007: 114), it can in fact be justified only if: (a) weights within strata are not correlated with survey variables, (b) relative variances of weights are equal in each stratum, and (c) the stratum intra-cluster correlation coefficients do not differ across the strata. Where one or more of these conditions does not exist, one must estimate the measure of variance inflation in some other way, as recently observed by Lee (2012: 16-20).

Figure 1 Assumption used in $DEFF_{TOTAL}$ estimation



Apart from these three criteria for the estimation of $DEFF_{TOTAL}$ as the product of the effects resulting from stratification, clustering and weighting, analyses of other conditioning criteria have appeared in the methodological literature and on this basis an additional criterion B1 may be pointed out, i.e., homogeneity of strata variances, which allows estimation of $DEFF_{TOTAL}$ without a stratification effect (Lee 2012: 19, Lynn et al. 2007: 114, Gabler et al. 2006: 115). In turn, Park et al. (2004: 183-193) focus on the consequences resulting from negating the condition described in criterion A1, which occur in those designs in which probabilities of selection are correlated with survey variables. Gabler et al. (2006: 116-117) investigate different variants of $DEFF_{TOTAL}$ estimation resulting from the adoption of criteria B1, B2, B2' and B3.

Furthermore, in a situation where clusters of different sizes are selected, there arises the problem of estimating parameter b'. Kish (1987) proposes using average cluster size,[4] in which case b' may be expressed as $\bar{b} = \frac{1}{m}\sum_{j=1}^{m} b_j$, where m denotes the number of clusters drawn and $b_1, b_2, \ldots, b_m$ denote the number of individuals in the j-th cluster. Such an estimation, however, proves risky: as observed by Gabler et al. (1999), using average cluster size in a cluster sampling design with unequal probabilities of selection of individuals may lead to a significant underestimation of DEFF$_c$. The authors argue that a more accurate estimation of DEFF$_c$ may be achieved by defining the parameter b' as an average weighted cluster size,[5] i.e., as $b^* = \frac{\sum_{j=1}^{m}\left(\sum_{i=1}^{l_j} w_{ji}\right)^2}{\sum_{j=1}^{m}\sum_{i=1}^{l_j} w_{ji}^2}$ (Gabler et al. 1999: 105), or in the form of a formula $\bar{b}_w = \frac{\sum_{j=1}^{m} l_j \sum_{i=1}^{l_j} w_{ji}^2}{\sum_{j=1}^{m}\sum_{i=1}^{l_j} w_{ji}^2}$ (Gabler et al. 1999: 105), where for each j-th cluster (or j-th set of respondents) with size equal to $b_1, b_2, \ldots, b_m$, the value w_{ji} stands for the weight given to the i-th individual from the j-th cluster.

THE PROBLEM

The subject of analysis in this paper will be the sampling scheme in the Polish section of round 5 of the European Social Survey from 2010 (ESS5-PL). In this project, stratified sampling is adopted: the population is divided into strata defined on the basis of the size of towns and villages. In the sample design related to the Polish part of ESS, it is important to note that in five strata, i.e., in 86 cities with at least 50,000 inhabitants, simple random sampling is used, with sample size proportional to population size, whereas cluster sampling is used in the village stratum and the three strata including cities with no more than 49,900 inhabitants. In the cluster part of a sample, sampling is done in three stages. The primary sampling units within each stratum are communities and places within their limits, selected with replacement and with a probability proportional to population size. Within the strata of villages and smaller cities, clusters involving 4 respondents are chosen. In the course of the sampling procedure, the probability of a unit from among the population being selected into a sample is also determined (*Sampling design in ESS5-PL* 2010: 2-3). The sampling scheme in ESS5-PL thus has the form of a multi-stage sampling, with a selection of individuals from stratified populations and partial clustering of the research sample. In other words, when estimating DEFF$_{\mathrm{TOTAL}}$, the effects of stratification, clustering and weighting should all be taken into consideration.

Empirical analyses of data from ESS5-PL will be used (1) to assess the level of intra-cluster homogeneity of individuals in the cluster sample design, (2) to verify the criteria for the method of DEFF$_{TOTAL}$ estimation described above and (3) to compare values of DEFF$_{TOTAL}$ obtained by means of different estimation methods.

With regard to the first goal, two questions will be addressed:

Q1: Does the recommendation of the methodological documentation of ESS to adopt a prior estimate of intra-cluster correlation coefficients on the level of $\rho = 0.02$ (Lynn et al. 2007: 114) correspond with empirical reality, i.e., the values of this coefficient in a sample?

Q2: Is it possible to single out any sets or types of questions from all the questions included in the ESS5 questionnaire for which the level of intra-cluster correlation is stronger or weaker? Where do these differences arise from, and is it possible to determine any regularities in this respect?

The second goal concerns the criteria conditioning the method of $DEFF_{TOTAL}$ estimation in the sample design in ESS5-PL:

Q3: Which of the criteria involving stratification, clustering and weighing are borne out by ESS5-PL?

Evaluation of these criteria will be used in the comparison of different methods of estimation of $DEFF_{TOTAL}$.

Q4: Does adopting different (alternative) estimation procedures for sampling design effect lead to significant differences in the values of the measures obtained? If so, which factors are responsible for the differences among the estimated $DEFF_{TOTAL}$ values?

Framing this question requires precise specification. It should be noted that the basic difficulty arising from the use of a complex sampling scheme such as ESS5-PL involves the need to estimate $DEFF_{TOTAL}$ on the basis of a measure which will be consistent with the sampling scheme and the criteria which this scheme fulfils. In addition, it has already been noted that in many cases simplified measures are adopted, viz. measures based on idealized assumptions concerning the characteristic of a sample and its method of selection. Framing the fourth question in a slightly different way, the scope of analysis can be defined more precisely:

Q4’: Does the use of methodologically unjustified $DEFF_{TOTAL}$ estimation procedures lead to a significant overestimation (or underestimation) of the real values of $DEFF_{TOTAL}$?

Specification of compared variants of $DEFF_{TOTAL}$ estimation

The first procedure for estimation of $DEFF_{TOTAL}$ is used in the ESS5-PL. This variant requires that assumptions (A1, A1’, A2, B1, B2, B2’) are fulfilled. The basic form of the estimator is:

(2) $$\text{DEFF}_{TOTAL} = \text{DEFF}_p \cdot \sum_{j=1}^{H} h'_j\, \text{DEFF}_c^j .$$

In each stratum h'_j denotes the unweighted proportion of responding individuals. In ESS5-PL stratified sampling is adopted, with a division of the sample into two parts: cluster and non-cluster. In the cluster part it is assumed that the coefficients ρ are equal across strata; then $\text{DEFF}_c^I = 1 + (b' - 1)\rho$, where parameter $b' = \bar{b}$. In turn, in the non-cluster part $\text{DEFF}_c^{II} = 1$. Therefore, the effect of clustering concerns only part of the sample. The formula used to estimate this effect takes the form of $\text{DEFF}_c = h'_I\, \text{DEFF}_c^I + h'_{II}$, where $h'_I = \sum_{j=1}^{4} h'_j$ denotes the sum of proportions of consecutive strata from the cluster part of the sample, and $h'_{II} = \sum_{j=5}^{9} h'_j$ denotes the sum of proportions of consecutive strata from the non-cluster part of the sample. In turn, $\text{DEFF}_{TOTAL} = \text{DEFF}_p \cdot \text{DEFF}_c$ (cf. Lynn et al. 2007: 114 and *Sampling design in ESS5-PL* 2010: 2).

In the second variant the number of non-proportionally weighted strata in the sample is taken into consideration in relation to the number of strata in the population, assuming equality of the coefficient ρ across all strata of the cluster part of the sample. This variant requires that criteria (A1, A2, B1, ~B2, ~B3) are fulfilled. The basic form of the estimator is

(3) $$\text{DEFF}_{TOTAL} = \sum_{j=1}^{H} \frac{\hat{h}_j^2}{h'_j}\, \text{DEFF}_p^j\, \text{DEFF}_c^j ,$$

where $\hat{h}_j = \sum_{i=1}^{nj} w_{ji} / \sum_{i=1}^{n} w_i$ is the estimate of a fraction of the j-th strata in the population estimated on the basis of weights, defined as inverse probability of inclusion of individuals (Lee 2012: 18). This variant differs from the first variant in two ways. First, DEFF_p is estimated separately for each stratum and multiplied by DEFF_c before combining the stratum estimate of $\text{DEFF}_{\text{TOTAL}}$. Secondly, the stratum weights are defined as the proportion of the design weights in the stratum and not the proportion of respondents. Furthermore, this approximation of $\text{DEFF}_{\text{TOTAL}}$ does not take into account the part of DEFF_p that results from variation in selection probabilities between strata. As a result, $\text{DEFF}_{\text{TOTAL}}$ may be underestimated by this procedure. However, this approximation is suggested in the literature in this form for sampling design with equal weights within each domain/stratum (cf. scenario 2 in Gabler et al. 2006: 116). For data from ESS5-PL, the division of a sample into two parts is made: cluster, i.e., four strata considered together ($\text{DEFF}_c^I = 1 + (b' - 1)\rho$), parametre $b' = b^*$, and non-cluster, i.e., five strata considered together, for which the clustering effect $\text{DEFF}_c^{II} = 1$.

In the third procedure the number of non-proportionally weighted strata in the sample is taken into consideration (in relation to the number of strata in the population), assuming diversity of the coefficient ρ across all strata of the cluster part of the sample. Applying this procedure requires that the assumptions

A1, ~A2, B1, ~B2, ~B3 are fulfilled. The basic form of the estimator is the same as in the second variant (equation (3)), so it differs from II only in that ρ is now allowed to vary between strata. For data from ESS5-PL, this involves a division of the sample into five parts, where four parts correspond to strata of the population with clustering of respondents in a sample and one part corresponds to the non-cluster part ($DEFF_c^j = 1 + (b'_j - 1)\rho_j$ is calculated for each strata and parameter $b' = b^*$).

The first of these variants, which complies with the method originally used by the National Coordinators group in Poland, takes the form of a relatively simple formula. The possibility of its use, however, is restricted by methodological requirements. As regards Question (3),it is worth noting that in the procedure described in variant I the effect of stratification is omitted, which means that this procedure is appropriate only if condition B1 is fulfilled, viz. the condition of homogeneity of strata variances.[6] Secondly, because $DEFF_{TOTAL}$ is originally calculated in ESS5-PL as a product of the effect of clustering and weighting of data, criteria A1, A1' and A2 all have to be satisfied. Thirdly, this procedure is strictly correct only when weights are equal in the whole set of data (criterion B2), and proportionate stratification sampling is carried out in a proportional manner (criterion B2'; cf. formula (6) and (9) in Gabler et al. 2006: 116).

Let us examine these conditions in the context of ESS5-PL data. Firstly, it can be noted that, in the ESS5-PL, omitting the effect of stratification leads to a slight overestimation of $DEFF_{TOTAL}$. Averaging values of $DEFF_s$ reveals that stratification ensures a slightly greater effectiveness than sampling from the whole population (the average value of $DEFF_s$ was set at 0.997). In other words, estimating $DEFF_{TOTAL}$ without the effect of stratification results in a slight overestimation of this measure.[7] Furthermore, in ESS5-PL, criteria A1 and A2 are fulfilled. Weights are constructed in such a way that within each cluster (and practically within each stratum) all the units are ascribed the same weight. Somewhat more problematic is criterion A2, which requires equality of intra-cluster correlation coefficients in successive strata of a population. It can be noted that coefficient ρ was set at the level of 0.15 in the whole set, while in successive strata the intra-cluster correlation coefficients were set at: (a) 0.168 in the village strata, (b) 0.161in cities with fewer than 10,000 inhabitants, (c) 0.133 in cities having between 10,000 and 19,999 inhabitants, (d) 0.093 in cities having between 20,000 and 49,999 inhabitants. Even though the differences may not be significant, estimating $DEFF_c$ by taking into consideration or omitting intra-cluster diversity of ρ may result in significant differences in $DEFF_{TOTAL}$, since this value is not only conditioned by the level of intra-cluster homogeneity of units, but also by the size of the clusters, the fraction of sample units from a given population stratum, and weights ascribed to successive respondents.

Let us move to the criteria for the estimation of $DEFF_{TOTAL}$ involving weighing of data, i.e., criteria B2, B2' and B3. Notice that the ESS research conducted in Poland clearly fails to fulfil two fundamental criteria conditioning the use of the first method of estimation of this measure, i.e., B2 and B2'. Our major reservation concerns the way in which the value of $DEFF_c$ is established, or to be more precise, the use of parameter h'_j, which represents the simple proportion of the units in a sample belonging to the j-th stratum of the population. Because in ESS5-PL the data set is weighted (criteria B2 and B2' have to be disregarded), and the weighted number of strata in the sample is not proportional to the number of strata in the population (criterion B3 is not met),weighted non-proportional structure of strata in the sample should be used in the estimation of $DEFF_{TOTAL}$ instead of a parameter h'_j. In other words, the method of estimation of $DEFF_{TOTAL}$ applied in ESS5-PL does not fully comply with the sampling scheme used. This poses the following question: does this procedure of estimation of $DEFF_{TOTAL}$ result in a significant underestimation or overestimation of its "real" value?

In the following analysis, the procedure described in the first variant will be compared with two methods of estimating $DEFF_{TOTAL}$ that are appropriate for the sampling scheme used in ESS5-PL (cf. formula (8) in Lee 2012: 18 and formula (5) in Gabler et al. 2006: 116).[8] The two procedures will be analysed with respect to the assumed values of intra-cluster correlation coefficients across strata. In variant II, I will assume the equality of such coefficients in consecutive strata of the population (criterion A2), whereas in variant III, I will assume variability of the value ρ across strata (contrary to criterion A2). Both of these methods are justified in relation to the Polish section, round 5 of the ESS research. They will be compared in order to determine whether taking into consideration strata variability (ρ) influences in a significant way the estimation of the values of $DEFF_c$ and $DEFF_{TOTAL}$.

DESCRIPTION OF DATA

Before moving on to the results of the analyses, it is necessary to present a set of variables that describe the key characteristics of the data set from ESS5-PL. Out of over 200 questions included in the ESS5 questionnaire, 90 items fulfilling two criteria were selected for the analyses.[9] Firstly, only those questions asked of all of the respondents were taken into account. Secondly, the set of questions was narrowed down to such variables whose form of measurement allowed characterization of their values by means of an arithmetic mean. This second condition was necessary to establish the values of intra-cluster coefficients on the basis of variance analysis procedures.

The most important characteristics of the ESS5-PL sample are given in Table 1. It should be observed that clustering of the research sample concerned slightly more than 61.5% of all the respondents (who in all constituted 1078 people), ascribed to 353 different clusters. The remaining 673 people were selected by means of individual sampling, for which $DEFF_c$=1.

Table 1 Characteristics of the ESS5-PL dataset (ed. 2010) and the description of parameters used in DEFF estimation

Parametr	Population strata: Village	City below 10 thous.	City 10 – 19 thous.	City 20 – 49 thous.	City 50 – 99 thous.	City 100 – 199 thous.	City 200 – 499 thous.	City 500 – 999 thous.	Warsaw	Total
Population share h_j	37.94	6.02	6.99	10.99	8.50	8.14	9.47	7.31	4.64	100.0
	(61.94)				(38.06)					
Estimated population share $\hat{h}_j$	41.58	6.62	7.22	10.82	7.51	7.43	8.61	6.04	4.17	100.0
	(66.24)				(33.76)					
Sample shares h'_j	36.61	6.68	7.34	10.91	7.59	8.74	10.11	7.08	4.94	100.0
	(61.54)				(38.46)					
Sample size n_j	641	117	129	191	133	153	177	124	86	1751
	(1078)				(673)					
Number of clusters b_j	219	38	38	58	-	-	-	-	-	353
Mean cluster size $\bar{b}_j$	2.927	3.079	3.395	3.293	-	-	-	-	-	3.054
Weighted cluster size b^*_j, $\bar{b}_{wj}$	3.234	3.547	4.473	4.047	-	-	-	-	-	3.507
Intra-cluster correlation coefficient $\hat{\rho}_j^{ANOVA}$	0.168	0.162	0.133	0.093	-	-	-	-	-	0.150
$DEFF_{p_j}$	1.000	1.000	1.000	1.003	1.001	1.000	1.000	1.000	1.000	1.014
	(1.005)				(1.005)					

Source: Calculations on the basis of data repository ESS5-PL and ESS5-SDDF-PL

With respect to the parameters describing the size of clusters of respondents, the arithmetic mean of clusters in the ESS5-PL data set was 3.054,[10] whereas mean weighted size (expressed by the parameters b^* and $\bar{b}_w$) was set at 3.507.[11] Even though it is not always the case that $b^* > \bar{b}$ and $b^* = \bar{b}_w$, these relations are not random in ESS5-PL but result from the selected sampling scheme. The relations between these parameters were analysed by P. Lynn et al. (2005: 101-104), who argued with reference to the results of the first round of the ESS1 from 2002that:

> "There are five nations where sample units were individuals selected with equal probabilities (within clusters) from population registers (BE [Belgium], DE [Germany], HU [Hungary], PL [Poland], SI [Slovenia]). In this case (8) [e.g. equal weights within cluster] (and, therefore, (10) [e.g. the means and the variances of the weights within clusters do not depend on the clusters]) holds strictly" (Lynn et al. 2005: 104).

What clearly results from this fact is that b^* has to be equal $\bar{b}_w$; however, it does not explain the inequality $b^* > \bar{b}$. Lynn et al. (2005: 104)also point out that if the weights within each cluster are equal and the sizes of the clusters are different (as in the Polish ESS sample), then b^* will be greater than $\bar{b}$, as long as covariance between the size of the clusters and their weighted size is greater or equal to zero.[12] Simple mathematical operations conducted on the data from ESS5-PL allow us to confirm that this condition is fulfilled,[13] and therefore $b^* > \bar{b}$. In sum, it should be noted that because $b^* = \bar{b}_w > \bar{b}$, an estimate of DEFF$_c$ on the basis of the parameter of arithmetic mean (viz. $\bar{b}$) – as was originally done in the Polish section of this project – will be lower than the value for weighted size (b^* or $\bar{b}_w$).Therefore, the use of parameter b^* instead of $\bar{b}$ would require a sample of a bigger size than those drawn so far from calculations based on an arithmetic mean. Although this would increase the cost of research, such a procedure would be more justified from a methodological point of view for sample design with unequal probabilities of selection (Gabler et al. 1999: 105-106).

Considering the values of DEFF$_p$ in Table 1, an additional regularity of the ESS5-PL sampling scheme can be noted. Namely, the values of weights are equal not only within clusters of respondents, but also across all strata of the population. An exception to this rule is found in two city strata (cities with 20,000-49,999 inhabitants and 50,000-99,999 inhabitants), in which DEFF$_p$ values were slightly higher than 1. This is significant, as weighting of data in ESS5-PL results in barely any increase of variance with relation to the measurement within population strata; it does, however, lead to the loss of precision of estimation in the whole research sample. This also means that ESS5-PL fulfils criterion A1', namely the criterion concerning the equality of relative variances of weights in each stratum of the population under study.

RESULTS

Intra-cluster homogeneity

In the documentation of round V of the ESS project[14] (*Sampling for the ESS Round V* 2010), as well as in the studies devoted to earlier rounds (*Sampling for the ESS Round I* 2002: 5-6*, Sampling for the ESS Round II* 2004: 61), an interesting recommendation can be found, which has important practical and methodological consequences. In the sections concerning $DEFF_c$ estimation, the following statement appears:

> ”If there is no available empirical evidence at all upon which to base an estimate of *roh* [rate of homogeneity], then we suggest that a value of 0.02 should be used” (*Sampling for the ESS Round V* 2010: 14).

This statement is significant, for even in the case of such a methodologically advanced project as ESS, not much is known about the consequences that arise from the clustering of a research sample. While it is true that the recommended value of intra-cluster correlation coefficients was established even before the first round of ESS in 2002 on the basis of the results obtained in earlier research (cf. Lynn et al. 2007: 114), such an a priori assumption involves significant risk and may result in the underestimation of cluster effects. Of course recommending the adoption of a prediction of $\rho = 0.02$ only works in cases where there is no empirical means of predicting the coefficient *roh*. If the country had taken part in ESS previously, national coordinators were encouraged to make an effort to estimate the values of ρ.[15] However, in many countries additional analyses were not undertaken and only project recommendations were made (Sawiński 2011: 2).[16] This issue is of a fundamental nature; however, it does not directly concern the Polish part of the ESS study. Importantly, before drawing a sample for the ESS5-PL research (ed. 2010) additional methodological studies were conducted based on the data from ESS4-PL (ed. 2008), which aimed at estimating the level of intra-cluster homogeneity of units in the ESS research. These studies demonstrated that the average value of this coefficient was 0.12 (*Sampling Design in ESS5-PL* 2010: 3), which was significantly higher than the value recommended in the project documentation.

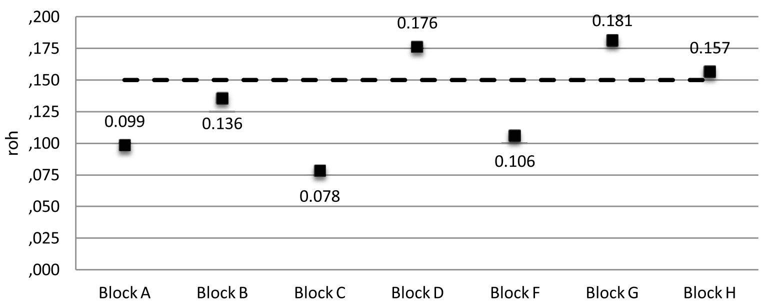
This discrepancy is confirmed by the analysis of the data from round V of ESS from 2010.The analysis reveals that the average value of the intra-cluster correlation coefficient was 0.15 (cf. the results in Table 1), which means that in ESS5-PL the level of intra-cluster homogeneity of units was slightly higher than even in ESS4-PL. These values, however, are not directly comparable for at least two reasons. Firstly, the questionnaires from the fourth and fifth round of the ESS study differ with respect to rotational modules (parts D and G in the research tools). Secondly, such longitudinal comparisons of general indicators are also limited by the fact that the analyses for the ESS4-PL study were conducted in relation to a significantly greater number of questions (183 variables) than the studies concerning ESS5-PL (90 variables) presented here. Despite these limitations, it is possible to compare the values of coefficients for particular questions in both rounds of ESS.

The results of methodological analyses devoted to the estimation of measurements of intra-cluster correlations in ESS4-PL were systematized by

Sawiński (2011: 1-16). The author focused on the similarity of units resulting from the clustering of a sample as well as the homogeneity of answers given by the respondents. The latter line of investigation was considered in the light of interviewer effect on the answers provided by the respondents. Such a duality of analysis is justified to some degree since in each selected cluster the interviews were conducted by the same interviewer, which means that similarity of units within clusters may result from cluster sampling as well as from interviewer effect. However, the method of conducting research applied inESS4-PLdid not allow one to estimating the increase in variance as a result of interviewer effect. Despite these limitations, the conclusions reached by Sawiński on the degree of interviewer effect are very valuable and are confirmed by ESS5-PL.

For example, in both rounds of ESS it turned out that a significantly lower level of intra-cluster homogeneity characterized those modules of the questionnaire which are placed at the beginning of the interview, whereas the questions asked at the end of the interview showed a higher level. Sawiński (2011: 4-5) explained this finding by means of the interviewer effect, which should be greater at the end of the questionnaire, when the respondents are more tired and prone to suggestions made by the interviewer. However, one would need a design with a randomized question order in order to be able to conclude that the difference is due to fatigue.

Figure 2 Values of intra-cluster correlation coefficients (roh) on the basis of the results from ESS5-PL (ed. 2010)



Source: Calculations on the basis of data repository ESS5-PL and ESS5-SDDF-PL

These correlations are also visible in ESS5-PL. The first three modules of the interview, namely A (media, trust in people), B (politics) and C (welfare, social exclusion, religion), which were included in both ESS4 and ESS5, are characterized by lower than average values of intra-cluster correlations. In turn, the rotational modules, namely D (trust in justice administration) and G (job, family, welfare), and the module of questions concerning human values (marked as H in ESS5 and as G in ESS4), are characterized by a higher level of homogeneity of responses given by respondents from the same clusters.

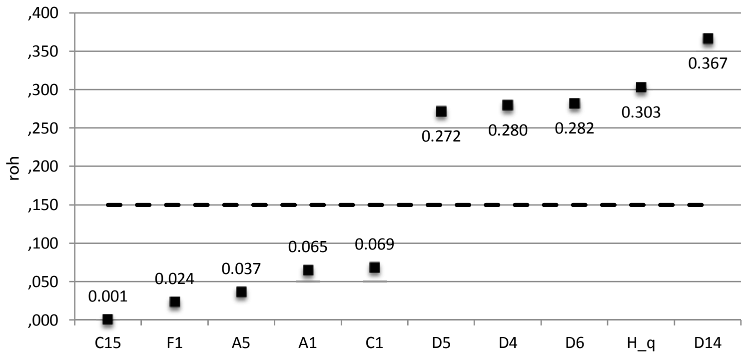
Although the hypothesis formulated by Sawiński concerning the relation between the value of ρ and the place of the question within the research tool seems plausible at first, further inspection raises certain problems. An analysis of values of the intra-cluster coefficients for ESS5-PL and ESS4-PL questions (cf. Sawiński 2011: 11-16) reveals that these values are actually quite variable within individual modules of the questionnaire. In addition, the highest level of intra-cluster homogeneity does not characterize questions from the last or next-to-last module of the questionnaire. Instead, it appears that the value ρ depends on the context of a question rather than its place in a research tool, i.e., the interview. Indeed, Sawiński came to such a conclusion in his investigation of the results of ESS4-PL:

> ”Questions that tackle problems which are important and controversial for the respondents usually have lower *roh* values […]. This is consistent with the view that higher respondent engagement results in a weaker interviewer effect” (Sawiński 2011: 5).

To support this argument, Sawiński quoted estimates of ρ values for the following questions: B1 (interest in politics; $\rho = 0.02$), B23 (place on the scale right-wing vs. left-wing; $\rho = 0.01$) and C15 (self-assessment of one's health; $\rho = 0.01$) and contrasted them with the much greater values of intra-cluster correlation coefficients for two groups of questions concerning the prestige of occupations.[17] In his assessment these questions concern issues that are not very important for the respondents, in contrast with issues discussed in questions B1, B23 and C15.

However, the assumption that important and controversial questions are associated with a lower level of intra-cluster homogeneity, and that general questions on less important issues are coupled with a higher level, is not confirmed by ESS5-PL research. When considering the data I observed that among the questions quoted by Sawiński from ESS4-PL characterized by a low level of intra-cluster homogeneity, the value of ρ was set at a similarly low level in ESS5-PLonly with relation to question C15.For the two remaining questions cited by Sawiński, significantly higher values of coefficients ρ were obtained in ESS5-PL (B1 – $\rho = 0.098$, whereas B23 – $\rho = 0.121$). In fact, among the five variables with the highest level of intra-cluster homogeneity in ESS5-PL three questions (D4, D5 and D6) were certainly “controversial” for the respondents and one question concerned an important issue for the respondents (question D14). This last variable was characterized by the highest level of intra-cluster homogeneity in the entire set of results from ESS5-PL.

Figure 3 Maximum and minimum values of intra-cluster correlation coefficients (roh) in the results from ESS5-PL research (ed. 2010)



Source: Calculations on the basis of ESS5-PL and ESS5-SDDF-PLdata repositories

Comparing the two series of ESS research from 2008 and 2010, a slightly different conjecture can be made about the relationship between the type of question and value of the intra-cluster correlation coefficient. Namely, it can be observed that within the set of questions relating to direct life experiences of the respondents, their day-to-day activities and facts about their lives, the coefficient ρ will be lower than for the measurement of opinions or beliefs concerning general issues. Indeed, the variables characterized by the lowest values of intra-cluster homogeneity in the ESS5-PL dataset concerned questions about self-assessment of one's health condition (C15), the number of people living in a household (F1), time per week devoted to reading newspapers (A5) and watching TV (A1), and the feeling of personal happiness (C1). On the other hand, the variables for which the highest values of were obtained related to questions about opinions and beliefs. What seems to be particularly important with respect to this hypothesis is the fact that all the questions that were characterized by the lowest level of intra-cluster homogeneity in ESS5-PL were also characterized by lower than average values of ρ in the ESS4-PL measurement.

Design effect of an ESS5-PL sampling scheme

The results discussed above demonstrate that the procedure of estimation of $DEFF_{TOTAL}$ applied by the team of Polish coordinators has to meet several strict conditions. In fact, these criteria were not met by the ESS5-PL sampling scheme, which calls into question the accuracy of the estimation of $DEFF_{TOTAL}$. Therefore, two alternative procedures for estimation of this measure have been proposed here, i.e., variants II and III, for which the formal criteria for use inESS5-PL were satisfied. The main purpose of this section will be to provide an answer to question 4, i.e., whether the use of alternative procedures leads to any significant differences in the values of $DEFF_{TOTAL}$.

Table 2 lists the average values of $DEFF_c$ and $DEFF_{TOTAL}$ provided jointly for successive modules of the questionnaire as well as their total value for all of the questions from ESS5.

Table 2 Comparison of the mean $DEFF_c$ and $DEFF_{TOTAL}$ values in selected data from ESS5-PL research (ed. 2010)

ESS5-2010	Estimation variant					
	$DEFF^{I}_{c}$	$DEFF^{II}_{c}$	$DEFF^{III}_{c}$	$DEFF^{I}_{TOTAL}$	$DEFF^{II}_{TOTAL}$	$DEFF^{III}_{TOTAL}$
Block A (6 items)	1.124	1.186	1.179	1.140	1.191	1.181
Block B (22 items)	1.172	1.252	1.256	1.188	1.258	1.257
Block C (5 items)	1.099	1.150	1.153	1.114	1.155	1.154
Block D (31 items)	1.223	1.325	1.334	1.240	1.331	1.336
Block F (2 items)	1.144	1.177	1.161	1.160	1.182	1.162
Block G (3 items)	1.229	1.334	1.336	1.246	1.340	1.337
Block H (21 items)	1.198	1.290	1.292	1.214	1.296	1.293
Total (90 items)	1.189	1.277	1.281	1.205	1.283	1.283

Source: Calculations on the basis of data repository ESS5-PL and ESS5-SDDF-PL

It should be noted that clustering has a much greater influence on the value of the total effect of sampling design than weighting of data that compensates for unequal probabilities of the selection of individuals for a research sample. In comparing the values of $DEFF_c$ and $DEFFF_{TOTAL}$ (estimated by each of the three procedures considered here), it can be observed that the effect of clustering constitutes approximately 99% of the total value of $DEFF_{TOTAL}$, regardless of the method of estimation adopted. In other words, unequal selection probability has an extremely marginal influence on the loss of precision of estimation for ESS5-PL, as already pointed out in studies devoted to earlier rounds of this project (Vehovar 2007: 343, Gabler et al. 2006: 116).

The question is which differences in the assessment of the effectiveness of sampling scheme result from distinct procedures of estimation of $DEFF_c$ and $DEFF_{TOTAL}$. First, we will compare the second and third methods of estimation of these coefficients to find out whether the diversity of intra-cluster homogeneity in strata influences in any way estimations of $DEFF_c$ and $DEFF_{TOTAL}$. A comparison of general values shows that the discrepancies are small for both measurements (0.006 for $DEFF_c$ and likewise for $DEFF_{TOTAL}$). This results from the small diversity of ρ coefficients in the population strata (cf. Table 2), as has already been highlighted in analyses devoted to intra-cluster homogeneity of units. In other words, the ESS5-PL sample is characterized by a significant discrepancy in the values of $DEFF_c$ as well as $DEFF_{TOTAL}$. If the differences across clusters in the values of intra-cluster correlation coefficients were greater, then variant III of estimation would give more accurate measurements than variant II.

Table 2 also shows that the simplified procedure of estimation of $DEFF_c$ and $DEFF_{TOTAL}$ results in a smaller increase in variance than the second and third method. It can be observed that the value of $DEFF_c$ in the first variant turned out to be c. 9 pp smaller than the value obtained in variants two and three. For $DEFF_{TOTAL}$ the difference was slightly smaller, i.e., c. 8 pp. This finding is extremely important, as it indicates that using the simplified version of $DEFF_c$ and $DEFF_{TOTAL}$ estimators may result in underestimation of the degree to which a particular sampling scheme influences the loss in estimation accuracy. On the other hand, if parameter $b' = b^*$ was adoptedinstead of parameter $b' = \bar{b}$ in variant I, similarly to variants II and III, then the values estimated by means of these three procedures would be quite similar. These minor differences would be a consequence of the low variability in selection probabilities and a great similarity of the structure of strata in the sample and in the general population. Suffice it to say that for ESS5-PL data, the degree of dissimilarity of empirical distribution of strata in the sample and the distribution of the same strata in the population, as measured by the index of dissimilarity,[18] equals only 2.5%. With such a similar structure of strata in a sample and a population as in the ESS5-PL study, ignoring the effect of disproportionate stratification would result in the underestimation of $DEFF_{TOTAL}$ by a multiple of this measure equal to 1.014, i.e., by 1.4%.[18]Of course, if these disproportions were larger, the differences in the values of these measurements would be more significant.

SUMMARY

It has been shown here that drawing a representative sample involves several methodological difficulties. The challenges lie not only in choosing the best strategy for selecting respondents but also in defining estimators of design effect which allow an accurate estimation of this measure. This is not a simple matter, mostly due to the fact that it is not possible to identify all the schemes according to which a sample may be selected; only some general categories may be identified in this respect (cf. Groves 1989: 253). In any case, it can be observed that the variants of estimating design effect examined in this article are appropriate for the ESS5-PL sampling scheme, but may not be appropriate in other sampling schemes. In other words, the purpose of the analyses presented here was not to work out universal solutions for sampling, but only to indicate some general difficulties and the need to conduct further research in this area.

NOTES

1 Estimation of the indicator (in generall $DEFF_{TOTAL}$) has application to the size of research samples. The term *effective sample size* is well-known (Biemer 2011: 232, Kohler 2007: 56, Lynn et al. 2007: 112, Gabler et al. 2006: 4, Groves et al. 2004: 108), and relates to the size of a sample selected according to a particular sampling scheme scaled to simple random sample of corresponding size. In other words, an effective sample size is a theoretical size of a simple random sample such that it allows estimation of values of parameters with the same level of statistical error as sample sizes selected according to the scheme set by a researcher (Dorofeev et al. 2006: 90). This means that research procedures that aim at establishing the smallest research samples possible require de facto making them real according to an effective size, otherwise the final level of statistical error may go beyond the set maximum values. Most often then, the demanded effective size of a simple random sample (n_{eff}) is established, the increase of variance scale for particular sampling scheme is calculated ($DEFF_{TOTAL}$) and the sample size n_{SAMPLE} is established, equal to the size of n_{eff}. An n_{SAMPLE} size may be established by means of a simple formula: $n_{SAMPLE} = n_{\mathrm{eff}} \times DEFF_{TOTAL}$.

2 See Lee (2012: 19), Biemer (2011: 214), Lynn et al. (2007: 107-124), Gabler et al. (2006: 115-116), Gabler et al. (1999: 105-107), Park et al. (2004: 4-14).

3 The measurement of intra-cluster correlation originally took the proper form for selecting clusters of equal size (Kish 1965:171), and as such it may often be found in the literature (Gabler et a 2008: 194, Barnett 1974). A formula that allows the computation of intra-cluster correlation coefficients in case of clusters of unequal size can be found in a study by Kendall and Stuart (1979). The between-cluster correlation coefficient does not take into consideration all pairs of observations (as in classical correlation coefficients), but only pairs of elements within clusters (Dorofeev et al. 2006: 95). The estimation of values of intra-cluster correlation coefficients is based on a variance analysis procedure (Gabler et al. 2008: 196, Ukoumunne 2002: 3760, Groves 1989: 363-364). It is important to note that the estimation of intra-cluster correlation coefficients by the *ANOVA* procedure is approximately unbiased, efficient, and a consistent estimator of ρ (Paul et. Al. 2003: 507-523). Such a method of estimation of intra-class correlation coefficients was used in the ESS research (Gabler et al. 2008: 197).

4 As a matter of fact, such a method of estimating parameter *b*', viz. by means of the arithmetic mean of size of clusters, is adopted in the Polish section of the ESS project (*Sampling design in ESS5-PL*). In the ESS, however, some inaccuracy in recommendations regarding the method of estimation of *b*' should be noted. Suffice it to say that parts of the documents in the ESS include a suggestion to base $DEFF_c$ estimation on average size of clusters (*Sampling for the European Social Survey Round V: Principles and*

Requirements 2010: 13), whereas other parts suggest using the coefficient $\bar{b}$ (Ganninger 2013: 4).

5 The analyses conducted by Gabler et al. (1999) were continued in a study by Lynn et al. (2005: 101-104). In this very interesting paper, focus was given to the mutual relations between the parameters $\bar{b}$, b^* and $\bar{b}_w$ parameters, as well as to the reliance of their values on cluster size and weighting procedures. On the basis of the analyses conducted by these authors, it can be shown that in relation to the Polish section of the ESS5-PL (ed. 2010) the following must $b^* > \bar{b}$ be and $\bar{b}_w$.

6 Levene's test was used in order to verify the hypothesis of equality of strata variance for successive variables included in the fifth round of the ESS. In can be seen from the data that the criterion of homogeneity of variance has to be rejected for 33 of the 90 questions under study.

7 Considering the detailed values of DEFF_s for each of the 90 questions under study it can be seen that the diversity of values of DEFF_s is quite low (for most of the questions asked of respondents, the level of effectiveness of stratification sampling is approximately 1). Stratification turned out to be most beneficial in relation to the measurement of income of the respondents, viz. in question F41 (DEFF_s=0.956), and least beneficial for the measurement of the variable concerning assessment of the speed of action taken by the police, viz. for question D14 (DEFF_s=1.031). In total, the value of DEFF_s was lower than 1 for 45 questions, or half of the variables under study, equal to 1 for 7 questions, and higher than 1 for 38 questions.

8 In the study of Gabler et al. (2006: 116-117), a clear suggestion may be found to estimate the effect of sampling design in the Polish section of the ESS on the basis of a procedure analogous to variant II or III.

9 The questions selected for the analysis include a considerable subset of the subjects covered in ESS5. Among the stable blocks, meaning those repeated in all successive rounds of the ESS, 6 questions were selected in module A, 22 in B, 5 in C, and 2 in F. In turn, in rotational modules, D, G and H the analysis covered 31, 3 and 21 questions respectively. All questions from module I were excluded (the set of questions that verify the correctness of answers given to previous questions), and also from module J, including questions addressed to the interviewers during the course of an interview). The set of variables, including their characteristics, is provided in table A.1. in the Appendix.

10 Within particular questions in ESS5-PL, the cluster sizes might differ from one another, which resulted from missing data and refusals to give answers to particular questions. This diversity was omitted in further parts of the analysis, and average values for the whole sample were accepted.

11 In ESS5-PL, it thus does not matter whether it will be b^*, or $\bar{b}_w$, as both of these parameters have equal values.

12 Lynn et al. provide a formula to establish the value of this covariance (Lynn et al. 2005: 101). Matching their proposition to the marks given here, this indicator may be expressed as a formula $\mathrm{Cov}(b_j, b_j\bar{w}_j^2) = \frac{1}{m}\sum_{j=1}^{m} b_j^2\bar{w}_j^2 - \frac{1}{m}\bar{b}\sum_{j=1}^{m} b_j\bar{w}_j^2$, where for each j-th cluster of respondents of size b_1, b_2, …, b_m as $\bar{b}$, denotes the arithmetic mean of all cluster sizes, while as $\bar{w}_j$ denotes the arithmetic mean of values of eights in the j-th cluster.

13 The covariance coefficient in the results of ESS5-PL is 1.609.

14 This refers to an instruction that is given to national coordinators of the research that describes sampling schemes (*Sampling for the ESS Round V* 2010).

15 In the ESS materials describing rules and requirements concerning sampling schemes, it was indicated that "[i]t is essential that National Coordinators and the fieldwork

organizations analyze the data from round I to V to calculate appropriate intraclass correlation coefficients for the sample designs used in their countries" (*Sampling for the ESS Round IV* 2012: 4, *Sampling for the ESS Round V* 2010: 3).

16 Sawiński (2011: 2) hypothesized that such a situation results mainly from the fact that "experiments are risky in this area because if a higher value of rho is found, the sample should be enlarged". He further suggests that although following recommendations included in the project documentation of the ESS may diverge from real values of intra-cluster coefficients in a research sample, they are adopted for practical reasons, to avoid an increase in costs connected with conducting a greater number of interviews.

17 The first group of variables is characterized by an average value of ρ=0.20 (22 items), whereas for the second group =0.22 (10 items) (Sawiński 2011: 5).

18 Index of dissimilarity is a commonly used measurement that enables comparison of two different distributions of the same variable. Its value range is 0 to 1: the closer the value is to zero, the more convergent are the distributions (Kuha et al. 2011: 376, Mulekar et al. 2008: 2099).

19 This value was established as a quotient of the values $DEFF^{II}_{TOTAL}$ and $DEFF^{I}_{TOTAL}$, omitting $DEFF_p$ and $DEFF_c$. In adopting these assumptions $DEFF^{I}_{TOTAL}$ =1, whereas $DEFF^{II}_{TOTAL}$=1,014.

REFERENCES

Barnett, Vic. 1982. *Elements of Sampling Theory*. Warszawa: Państwowe Wydawnictwo Naukowe.

Biemer, Paul P. 2011 *Latent Class Analysis of Survey Error*. New Jersey: John Wiley & Sons, Inc.

Dorofeev, Sergey and Peter Grant. 2006. *Statistics for Real-Life Sample Surveys. Non-Simple-Random Samples and Weighted Data*. Cambridge: Cambridge University Press

Gabler, Siegfried, Matthias Ganninger, Sabine Häder, and Ralf Munnich. 2008. "Design effect (deff)."In: *Encyclopaedia of Survey Research Methods*, edited by Paul J. Lavrakas. PLACE: SAGE Publications Inc., pp. 193-197.

Gabler, Siegfried, Sabine Häder and Partha Lahiri. 1999. "A model based justification of Kish's formula for design effects for weighting and clustering." *Survey Methodology* Vol. 25(1): 105-106.

Gabler, Siegfried, Sabine Häder and Peter Lynn. 2006. "Design effects for multiple design samples."*Survey Methodology* Vol. 32(1): 115-120.

Ganninger, Mathias. 2013. *The ESS Sample Design Data File (SDDF)*. Documentation of the European Social Survey.

Groves, Robert M.1989. *Survey Errors and Survey Costs*. New Jersey: John Wiley & Sons, Inc.

Groves, Robert M., Floyd J. Fowler, Mick P. Couper, James M. Lepkowski, Elanor Singer and Roger Tourangeau. 2004. *Survey Methodology*. New Jersey: John Wiley & Sons, Inc.

Kendall, Maurice G. and Alan Stuart. 1979. *The Advanced Theory of Statistics*. Vol. 2: *Inference and Relationship*. 4th ed. London: Griffin.

Kish, Leslie. 1965. *Survey Sampling*. New Jersey: John Wiley & Sons, Inc.

Kish, Leslie. 1987. *Statistical Design for Research*. New Jersey: John Wiley & Sons, Inc.

Kish, Leslie and Martin R. Frankel. 1974. "Inference from complex samples."*Journal of the Royal Statistical Society. Series B (Methodological)* Vol. 36(1): 1-37.

Kohler, Urlich. 2007. "Survey from inside: an assessment of unit nonresponse bias with internal criteria."*Survey Research Methods* Vol. 2(1): 55-67.

Kuha, Jouni and David Firth. 2011. "On the index of dissimilarity for lack of fit in loglinear and log-multiplicative models."*Computational Statistics and Data Analysis* Vol. 55(1):375-388.

Lee, Hyunshik. 2012. "How should one find out the contributions to the design effect (variance) made by each of the design components (stratification, clustering, weighting) of a complex sample design?"*Survey Statistician* Vol. 66: 16-20.

Lynn, Peter and Siegfried Gabler. 2005. "Approximation of b^* in the prediction of design effects due to clustering."*Survey Methodology* Vol. 31(1): 101-104.

Lynn, Peter, Siegfried Gabler, Sabine Häder and Seppo Laaksonen. 2007. "Methods for achieving equivalence of samples in cross-national surveys."*Journal of Official Statistics* Vol. 27(1): 107-124.

Mulekar, Madhuri S., John C. Knutson and Jyoti A. Champanerkar. 2008. "How useful areapproximations to mean and variance of the index of dissimilarity?" *Computational Statistics & Data Analysis* Vol. 52(4): 2098-2109.

Park, Inho and Hyunshik Lee. 2004. "Design effects for the weighted mean and total estimators under complex survey sampling."*Survey Methodology* Vol. 30(2): 183-193.

Paul, Sudhir R, Krishna K. Saha and Uditha Balasooriya. 2003. "An empirical investigation of different operating characteristics of several estimators of the intraclass correlation in the analysis of binary data."*Journal of Statistical Computation & Simulation* Vol. 73(7): 507-523.

Sampling design in ESS5-PL. 2010. Warszawa: Ośrodek Realizacji Badań Społecznych IFiS PAN.

Sampling for the European Social Survey Round I. 2002. Bergen: European Social Data Archive, Norwegian Social Science Data Service.

Sampling for the European Social Survey Round II. 2004. Bergen: European Social Data Archive, Norwegian Social Science Data Service.

Sampling for the European Social Survey Round V: Principles and Requirements. 2010. Mannheim: The Sampling Expert Panel of the ESS. GESIS.

Sawiński, Zbigniew. 2011. "Intra-cluster homogeneity in survey samples: a neglected tool."Paper presented at the 4th Conference of the European Survey Research Association (ESRA), Lausanne, 18-22 July 2011.

Ukoumunne, Obioha C. 2002. "A comparison of confidence interval methods for the intraclass correlation coefficient in cluster randomized trials."*Statistics in Medicine* Vol. 21(24): 3757-3774.

Vehovar, Vasja. 2007. "Non-response bias in the European Social Survey."In: *Measuring Meaningful Data in Social Research*, edited by G. Loosveldt, M. Swyngedouw, and B. Cambr. Leuven: ACCO, pp. 335–356.

Piotr Jabkowski is an Assistant Professor at the Institute of Sociology of Adam Mickiewicz University in Poznań. After obtaining an M.A. in applied mathematics at the Faculty of Mathematics and Computer Science of Adam Mickiewicz University in Poznań (2001), he began his doctoral studies at the Faculty of Social Sciences of Adam Mickiewicz University in Poznań, where he completed his Ph.D. in sociology in 2006. His research interests include problems related to the quality of data obtained in survey studies, including the total survey error paradigm and sampling theory.

E-mail: pjabko@amu.edu.pl

APPENDIX

Table A.1 Items selected from the ESS5 questionnaire

Question ID	Description	Values
	Module A – media, trust in people	
A1	TV watching, total time on average weekday	0 – 7
A3	Radio listening, total time on average weekday	0 – 7
A5	Newspaper reading, total time on average weekday	0 – 7
A8	Most people can be trusted or you can't be too careful	0 – 10
A9	Most people try to take advantage of you, or try to be fair	0 – 10
A10	Most of the time people helpful or mostly looking out for themselves	0 – 10
	Module B – politics	
B1	How interested in politics	1 – 4
B4	Trust in country's parliament	0 – 10
B5	Trust in the legal system	0 – 10
B6	Trust in the police	0 – 10
B7	Trust in politicians	0 – 10
B8	Trust in political parties	0 – 10
B9	Trust in the European Parliament	0 – 10
B10	Trust in the United Nations	0 – 10
B23	Placement on left right scale	0 – 10
B24	How satisfied with life as a whole	0 – 10
B25	How satisfied with present state of economy in country	0 – 10
B26	How satisfied with the national government	0 – 10
B27	How satisfied with the way democracy works in country	0 – 10
B28	State of education in country nowadays	0 – 10
B29	State of health services in country nowadays	0 – 10
B30	Government should reduce differences in income levels	1 – 5
B31	Gays and lesbians free to live life as they wish	1 – 5

B32	Ban political parties that wish overthrow democracy	1 – 5
B33	Modern science can be relied on to solve environmental problems	1 – 5
B38	Immigration bad or good for country’s economy	0 – 10
B39	Country’s cultural life undermined or enriched by immigrants	0 – 10
B40	Immigrants make country worse or better place to live	0 – 10
	Module C – feeling of welfare, social exclusion, religion	
C1	How happy are you	0 – 10
C4	Take part in social activities compared to others of same age	1 – 5
C6	Feeling of safety of walking alone in local area after dark	1 – 4
C15	Subjective general health	1 – 5
C21	How religious are you	0 – 10
	Module D – trust in justice administration	
D1	How wrong to make exaggerated or false insurance claim	1 – 4
D2	How wrong to buy something that might be stolen	1 – 4
D3	How wrong to commit traffic offence	1 – 4
D4	How likely be caught if made exaggerated or false insurance claim	1 – 4
D5	How likely to be caught if bought something that might be stolen	1 – 4
D6	How likely to be caught if committed traffic offence	1 – 4
D7	Police doing good or bad job in country	1 – 5
D12	How successful police are at preventing crimes in country	0 – 10
D13	How successful police are at catching house burglars in country	0 – 10
D14	How quickly would police arrive at a violent crime/burglary scene near to where you live	0 – 10
D18	Duty to: back decisions made by police, even if disagree	0 – 10
D19	Duty to: do what police say, even when don’t understand or agree	0 – 10
D20	Duty to: do what police say even if treated badly	0 – 10
D21	Police have the same sense of right and wrong as me	1 – 5
D22	Police stand up for values that are important to people like me	1 – 5
D23	I generally support how the police act	1 – 5
D24	Decisions and actions of police unduly influenced by political pressure	1 – 5
D25	How often do police in country take bribes	0 – 10
D26	Courts doing good or bad job in country	1 – 5
D27	How often the courts make mistakes that let guilty people go free	0 – 10
D28	How often the courts make fair, impartial decisions based on available evidence	0 – 10
D31	How often judges in country take bribes	1 – 5
D32	Courts protect rich and powerful over ordinary people	1 – 5
D33	People who break the law much harsher sentences	1 – 5

D34	Everyone's duty to back the court's final verdict	1 – 5
D35	All laws should be strictly obeyed	1 – 5
D36	Doing the right thing sometimes means breaking the law	1 – 5
D37	The courts' decisions are unduly influenced by political pressure	1 – 5
D40	How likely to call police if you see a man get his wallet stolen	1 – 4
D41	How willing to identify person who had done it	1 – 4
D42	How willing to give evidence in court against the accused	1 – 4
	Module F – sociodemographic characteristics	
F1	Number of people living regularly as member of household	
F41	Household's total net income, all sources	1 – 10
	Module G – Job, family, welfare	
G4	Women should be prepared to cut down on paid work for sake of family	1 – 5
G5	Men should have more right to job than women when jobs are scarce	1 – 5
G6	Government do more to prevent people falling into poverty	1 – 5
	Module H – human values	
H_a	Important to think new ideas and being creative	1 – 6
H_b	Important to be rich, have money and expensive things	1 – 6
H_c	Important that people are treated equally and have equal opportunities	1 – 6
H_d	Important to show abilities and be admired	1 – 6
H_e	Important to live in secure and safe surroundings	1 – 6
H_f	Important to try new and different things in life	1 – 6
H_g	Important to do what is told and follow rules	1 – 6
H_h	Important to understand different people	1 – 6
H_i	Important to be humble and modest, not draw attention	1 – 6
H_j	Important to have a good time	1 – 6
H_k	Important to make own decisions and be free	1 – 6
H_l	Important to help people and care for others well-being	1 – 6
H_m	Important to be successful and that people recognise achievements	1 – 6
H_n	Important that government is strong and ensures safety	1 – 6
H_o	Important to seek adventures and have an exciting life	1 – 6
H_p	Important to behave properly	1 – 6
H_q	Important to get respect from others	1 – 6
H_r	Important to be loyal to friends and devote to people close	1 – 6
H_s	Important to care for nature and environment	1 – 6
H_t	Important to follow traditions and customs	1 – 6
H_u	Important to seek fun and things that give pleasure	1 – 6